

Uses of Information Theory in Medical Imaging

Norbert Schuff, Ph.D.

Center for Imaging of Neurodegenerative Diseases

Norbert.schuff@ucsf.edu

With contributions from Dr. Wang Zhang

Topics

- I. Measurement of information
- II. Image registration using information theory
- III. Imaging feature selection using information theory
- IV. Image classification based on information theoretic measures

Measurement & Information

Objective

Learn how to quantify information

Information and Uncertainty

Random
Generator



ABACCACB

Which char
comes
next?

A

Decrease in
Uncertainty

$$\log(1/3) = -\log(3)$$



011000101

1

$$\log(1/2) = -\log(2)$$

Both combined:

$$-\log(3) - \log(2) = -\log(6)$$

Note, we assumed each symbol is likely to appear at equal chance

More On Uncertainty

Random
Generator
Of M Symbols



→ AB@DCE\$

Which one
comes
next?



Decrease in
Uncertainty

$$\log(1/M) = -\log(p)$$

Some symbols may appear more likely than others

$$-p_i \log_2(p_i)$$

$$\text{Many : } H_{\text{Shannon}} = -\sum_i p_i \log_2(p_i); \dots \sum_i p_i = 1$$

Example: Entropy Of mRNA

...ACGTAAACACCGCACCTG

$$p_A = \frac{1}{2}; p_C = \frac{1}{4}; p_G = \frac{1}{8}; p_T = \frac{1}{8}$$

$$\text{Bin} : -\log_2(p_i) : p_A = 1; p_C = 2; p_G = 3; p_T = 3$$

$$H = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1.75 \quad \text{Bits per symbol}$$

Concept Of Entropy

- **Shannon Entropy formulated by Claude Shannon**

- American mathematician
- Founded information theory with one landmark paper in 1948
- Worked with Alan Turing on cryptography during WWII



Claude Shannon
1916-2001

- **History of Entropy**

- **1854** – Rudolf Clausius, German physicist, developed theory of heat
- **1876** – Williard Gibbs, American physicist, used it for theory of energy
- **1877** – Ludwig Boltzmann, Austrian physicist, formulated theory of thermodynamics
- **1879** – Gibbs re-formulated entropy in terms of statistical mechanics
- **1948** - Shannon

Three Interpretations of Entropy

- The uncertainty in the outcome of an event
 - Systems with one common event have less entropy than systems with various common events.
- The amount of information an event provides
 - An infrequently occurring event provides more information. i.e. has higher entropy, than a frequently occurring event
- The dispersion in the probability distribution
 - A uniform image has a less dispersed histogram and thus lower entropy than a heterogeneous image.

Generalized Entropy

- The following generating function can be used as an abstract definition of entropy:

$$H(P) = h \left(\frac{\sum_{i=1}^M v_i \cdot \varphi_1(p_i)}{\sum_{i=1}^M v_i \cdot \varphi_2(p_i)} \right)$$

- Various definitions of these parameters provide different definitions of entropy.
 - Found over 20 definitions of entropy

Various Formulations Of Entropy

Names		$h(x)$	$\varphi_1(x)$	$\varphi_2(x)$	v_i
Shannon		x	$-x \log x$	x	v
Renyi		$(1-r)^{-1} \log x$	x^r	x	v
Aczel		x	$-x^r \log x$	x^r	v
Aczel		$(s-r)^{-1} \log x$	x^r	x^s	v
Aczel		$(1/s) \arctan x$	$x^r \sin(s \log x)$	$x^r \cos(s \log x)$	v
Varma		$(m-r)^{-1} \log x$	x^{r-m+1}	x	v
Varma		$(m(m-r))^{-1} \log x$	$x^{r/m}$	x	v
Kapur		$(1-t)^{-1} \log x$	x^{t+s-1}	x^s	v
Hadra		$(1-s)^{-1}(x-1)$	x^s	x	v
Arimoto		$(t-1)^{-1}(x^t-1)$	$x^{1/t}$	x	v
Sharma		$(1-s)^{-1}(e^x-1)$	$(s-1)x \log x$	x	v
Sharma		$(1-s)^{-1}(x^{\frac{s-1}{r-1}}-1)$	x^r	x	v

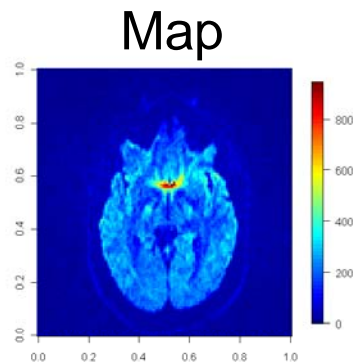
Various Formulations Of Entropy II

Names		$h(x)$	$\varphi_1(x)$	$\varphi_2(x)$	v_i
Taneja		x	$-x^r \log x$	x	v
Sharma		$(s-r)^{-1}x$	$x^r - x^s$	x	v
Sharma		$(\sin s)^{-1}x$	$-x^r \sin(s \log x)$	x	v
Ferreri	...	$\left(1 + \frac{1}{\lambda}\right) \log(1 + \lambda) - \frac{x}{\lambda}$	$(1 + \lambda x) \log(1 + \lambda x)$	x	v
Sant'Anna		x	$-x \log \left(\frac{\sin(sx)}{2 \sin(s/2)} \right)$	x	v
Sant'Anna		x	$-\frac{\sin(xs)}{2 \sin(s/2)} \log \left(\frac{\sin(sx)}{2 \sin(s/2)} \right)$	x	v
Picard		x	$-x \log x$	x	w_i
Picard		x	$-\log x$	1	v_i
Picard		$(1-r)^{-1} \log x$	x^{r-1}	1	v_i
Picard		$(1-s)^{-1}(e^x - 1)$	$(s-1) \log x$	1	v_i
Picard		$(1-s)^{-1}(x^{\frac{r-1}{s-1}} - 1)$	x^{r-1}	1	v_i

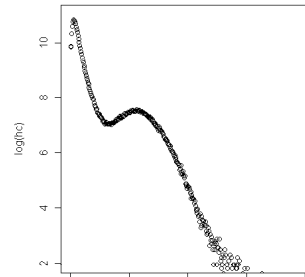
Entropy In Medical Imaging

Averaging

N=1



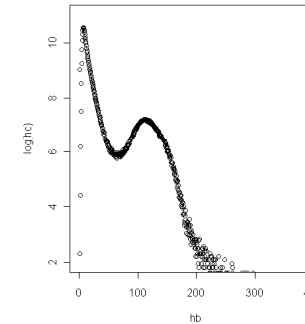
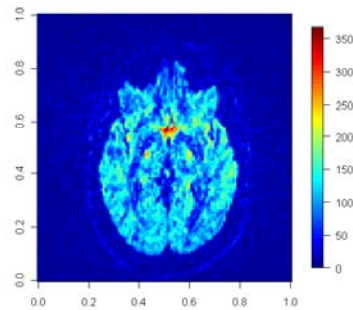
Histogram



Entropy

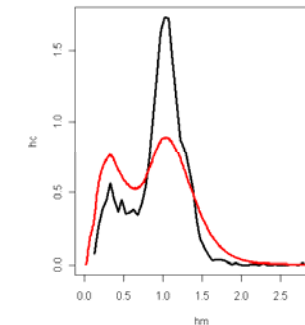
H=6.0

N=40



H=4.0

Histograms
image brain only
Co-variance map



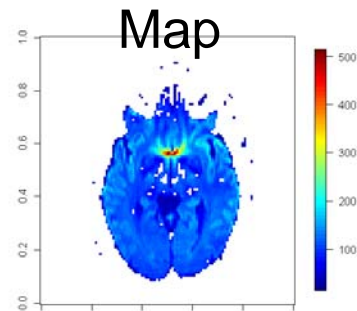
H=3.9

H=10.1

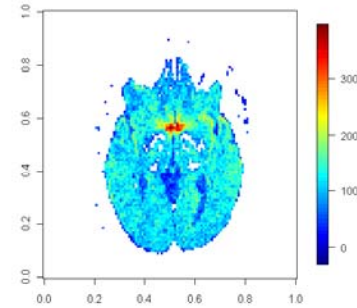
Entropy Of Noisy Images

Signal-to-Noise
Level

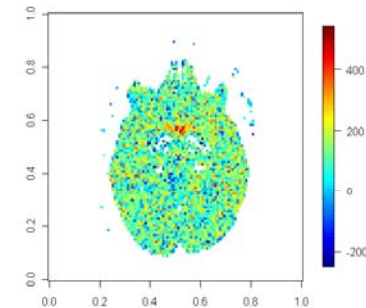
SNR=20



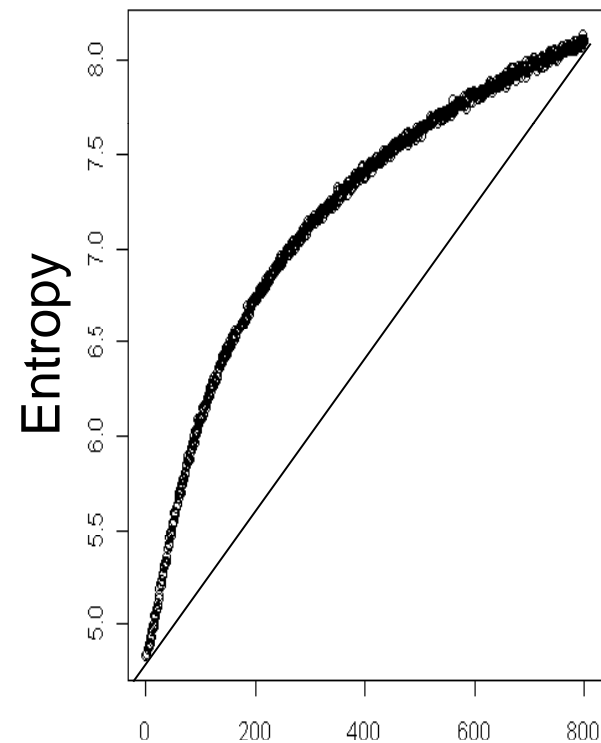
SNR=2



SNR=0.2



Entropy



Decreasing SNR Levels

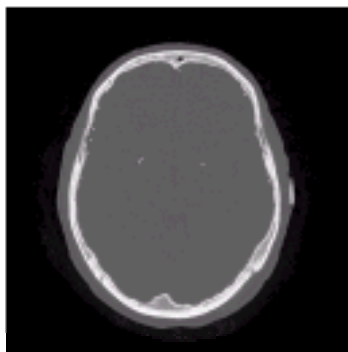
Image Registration Using Information Theory

Objective

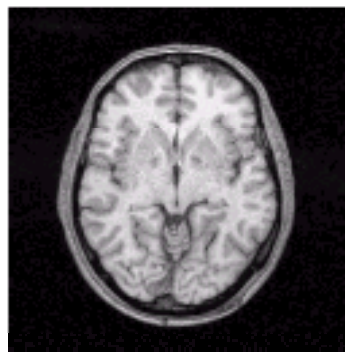
Learn how to use Information Theory for
imaging registration and similar
procedures

Image Registration

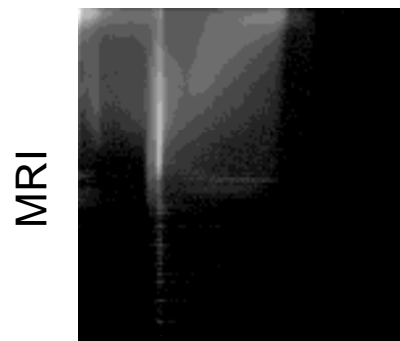
- Define a transform T that maps one image onto another image such that some measure of overlap is maximized (Colin's lecture).
 - Discuss information theory as means for generating measures to be maximized over sets of transforms



CT



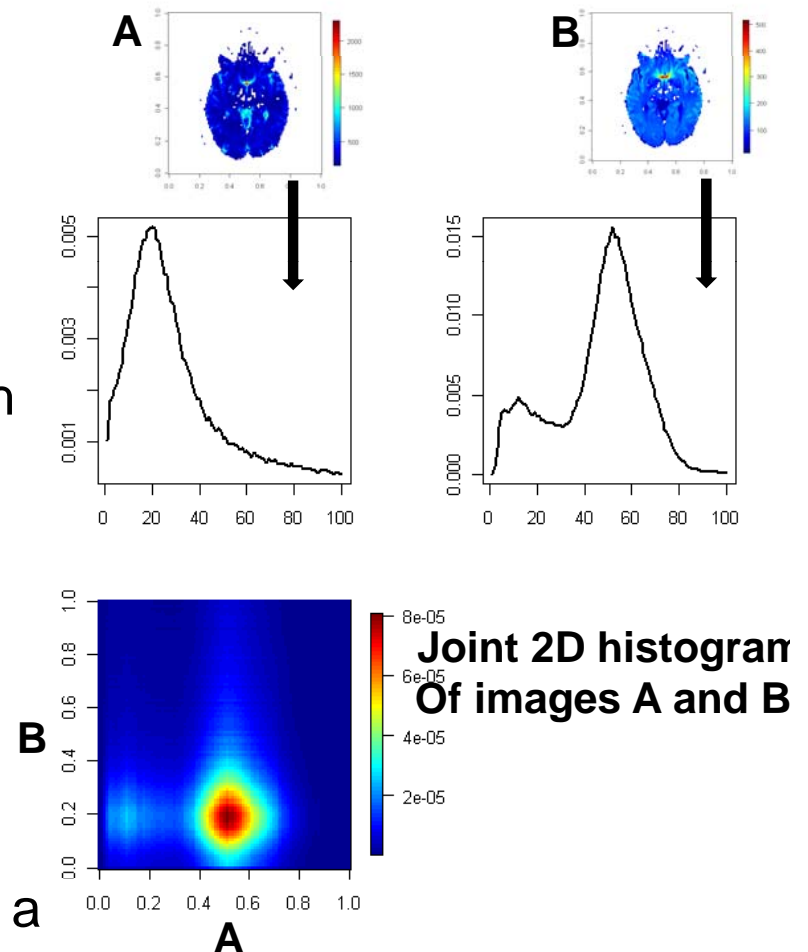
MRI



CT

Entropy In Image Registration

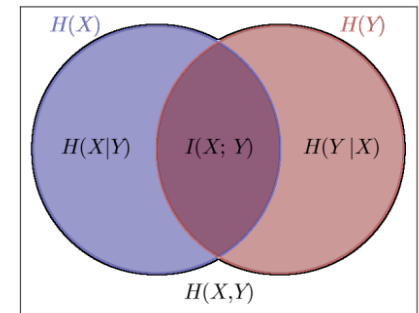
- Define estimate of joint probability distribution of images:
 - 2-D histogram where each axis designates the number of possible intensity values in corresponding image
 - each histogram cell is incremented each time a pair $(I_1(x,y), I_2(x,y))$ occurs in the pair of images (“co-occurrence”)
 - if images are perfectly aligned then the histogram is highly focused; as the images become mis-aligned the dispersion grows
 - recall one interpretation of entropy is as a measure of histogram dispersion



Use Of Entropy for Image Registration

- Joint entropy (entropy of 2-D histogram):

$$H(X, Y) = - \sum_{x_i \in X, y_i \in Y} p(x_i, y_i) \cdot \log_2[p(x_i, y_i)]$$



- $H(X, Y) = H(X) + H(Y)$ only if X and Y are completely independent.
- Image registration can be guided by minimizing joint entropy $H(X, Y)$, i.e. dispersion in the joint histogram for images is minimized

Example

Joint Entropy of 2-D Histogram for rotation of image with respect to itself of 0, 2, 5, and 10 degrees



3.82



6.79



6.98

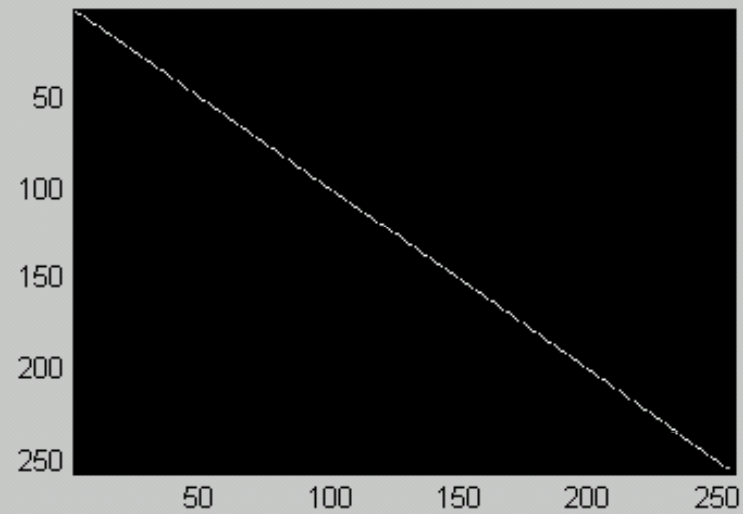


7.15

reference image



joint entropy = 5.53 M.I = 5.53 $[I(A,A)=H(A)]$



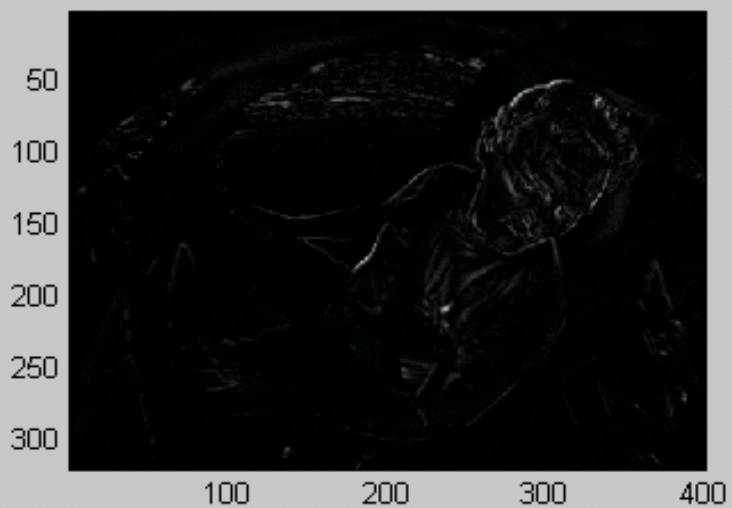
reference image



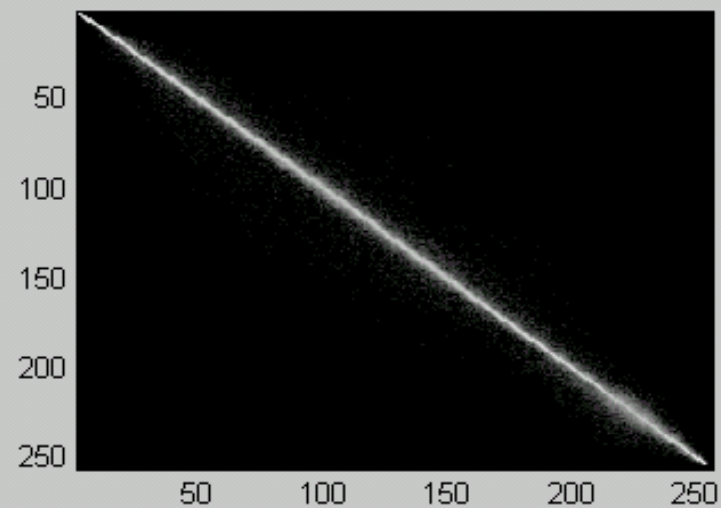
current image



difference image



joint entropy = 7.48 M.I.= 3.59



reference image



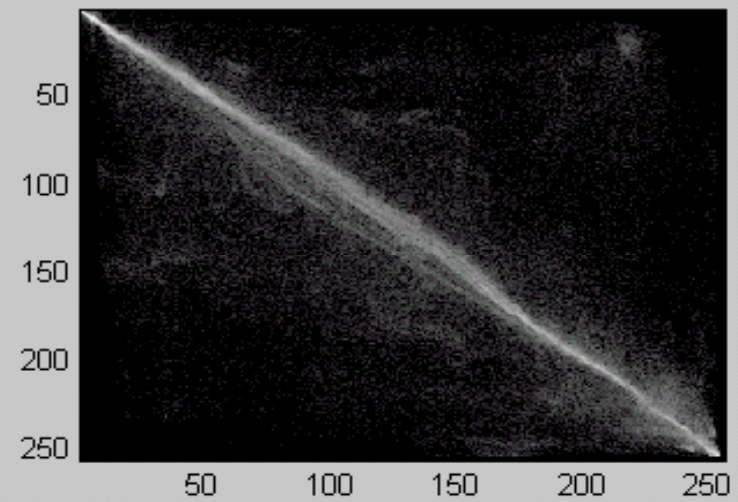
current image



difference image



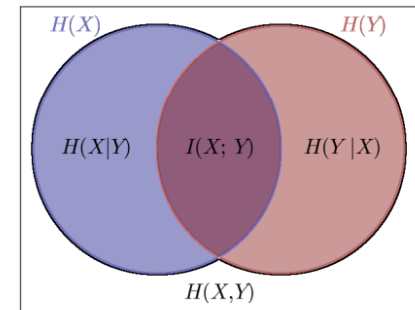
joint entropy = 9.36 M.I. = 1.70



Alternative: Conditional Entropy

- Assume we know X then we seek the remaining entropy of Y

$$\begin{aligned} H(Y | X) &= - \sum_{x_i \in X} p(x_i) \sum_{y_i \in Y} p(y_i | x_i) \cdot \log[p(y_i | x_i)] \\ &= - \sum_{x_i \in X, y_i \in Y} p(x_i, y_i) \log[p(y_i | x_i)] \\ &= - \sum_{x_i \in X, y_i \in Y} p(x_i, y_i) \log \left[\frac{p(y_i, x_i)}{p(x_i)} \right] \end{aligned}$$



Use Of Mutual Information for Image Registration

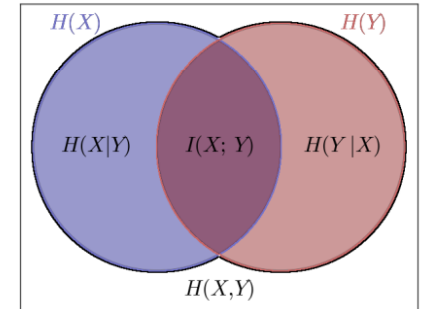
- Recall definition(s):

- $I(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$

- amount by which uncertainty of X is reduced if Y is known.

- $I(X,Y) = H(X) + H(Y) - H(X,Y)$

- maximizing $I(X,Y)$ is equivalent to minimizing joint entropy $H(X,Y)$



- Advantage in using mutual information (MI) over joint entropy is that MI includes the entropy of each distribution separately.
- MI works better than simply joint entropy in regions with low contrast where there will be high joint entropy but this is offset by high individual entropies as well - so the overall mutual information will be low
- Mutual information is maximized for registered images

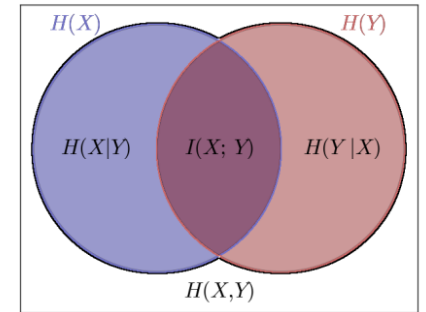
Formulation Of Mutual Information In Terms Of Entropy

$$I(X,Y) = \sum_{x_i \in X, y_i \in Y} p(x_i, y_i) \cdot \log \left(\frac{p(x_i, y_i)}{p(x_i)p(y_i)} \right)$$

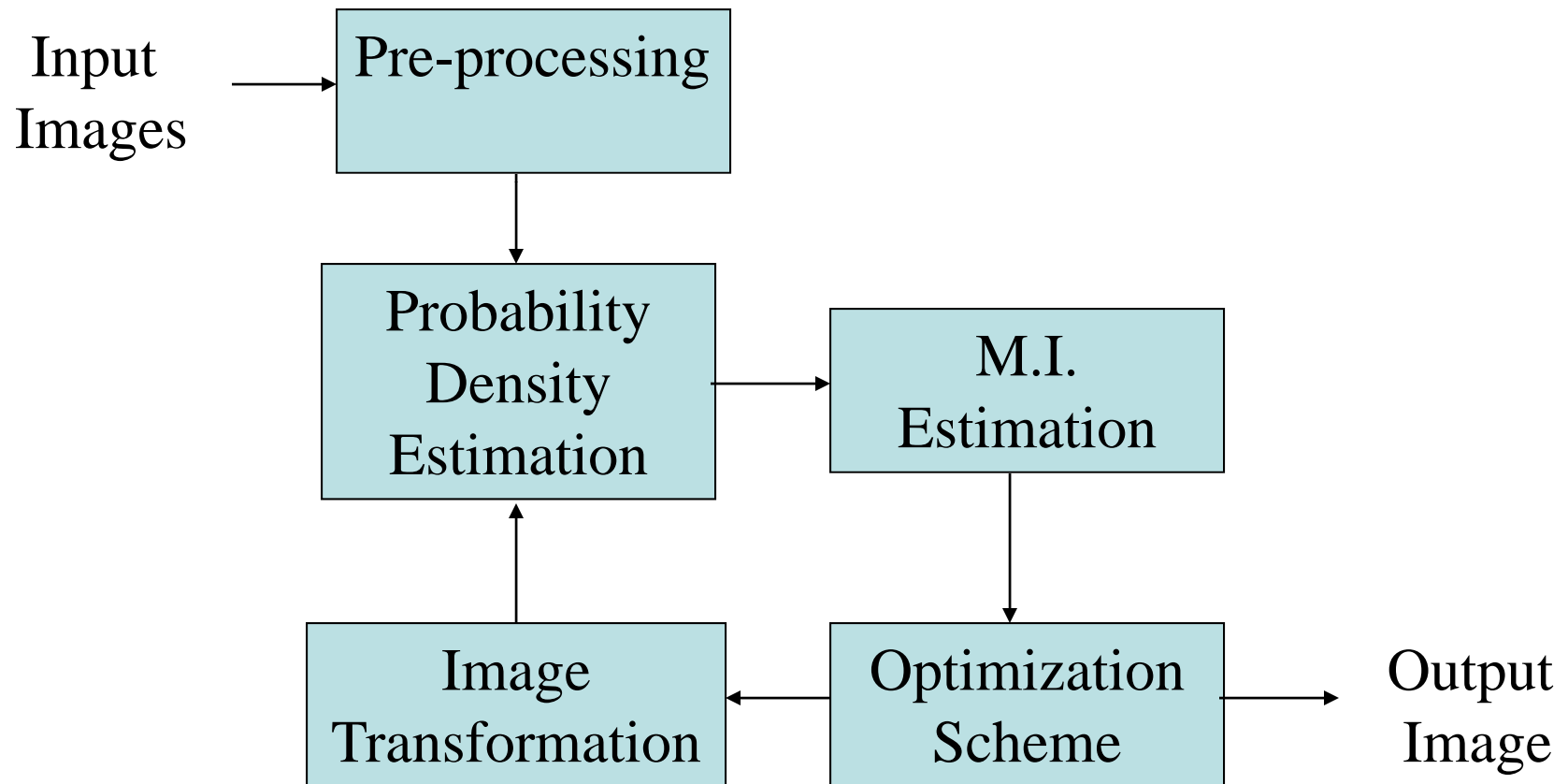
- Measures the dependence of the two distributions
- Is related to relative entropy (will be discussed next)
- In image registration $I(X,Y)$ is maximized when the images are aligned
- In feature selection choose the features that minimize $I(X,Y)$ to ensure they are not related.

Properties of Mutual Information

- MI is symmetric: $I(X, Y) = I(Y, X)$
- $I(X, X) = H(X)$
- $I(X, Y) \leq H(X)$
 - Information each image contains about the other image cannot be greater than the total information in each image.
- $I(X, Y) \geq 0$
 - Cannot increase uncertainty in X by knowing Y
- $I(X, Y) = 0$ only if X and Y are independent



Processing Flow for Image Registration Using M.I.



Measurement Of Similarity Between Distributions

- Question: How close (in bits) is a distribution X to a model distribution Ω ?

$$D(X \parallel \Omega)_{KL} = \sum_{x_i \in \Omega} p(x_i) \cdot \log \left(\frac{p(x_i)}{q(x_i)} \right)$$

Kullback-Leibler Divergence

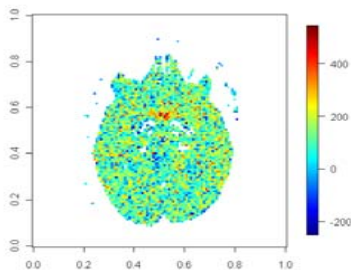
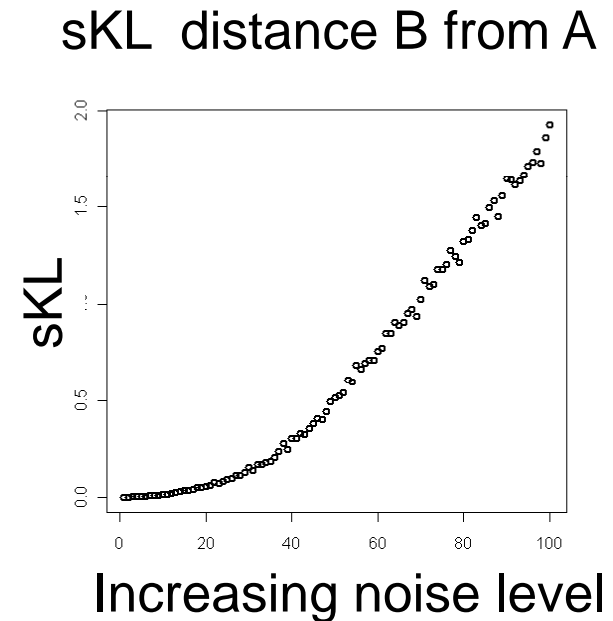
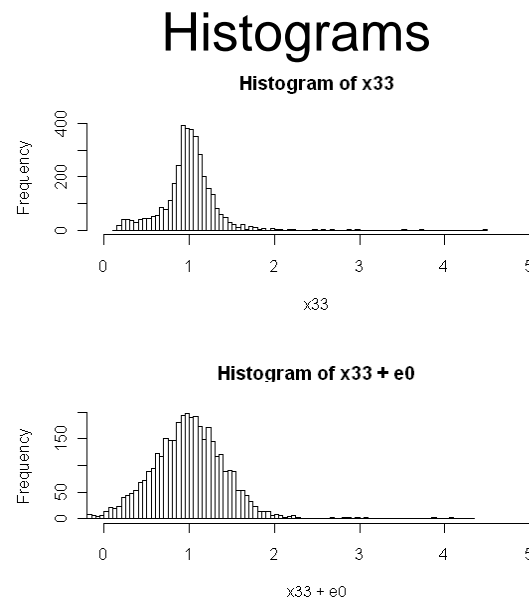
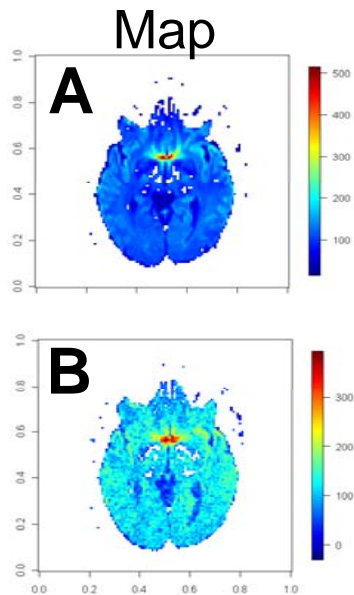
- $D_{KL} = 0$ only if X and Ω are identical; otherwise $D_{KL} > 0$
- $D_{KL}(X \parallel \Omega)$ is NOT equal to $D_{KL}(\Omega \parallel X)$

- Symmetrical Form

$$sD(X \parallel \Omega)_{sKL} = \frac{1}{2} (D(X \parallel \Omega) + D(\Omega \parallel X))$$

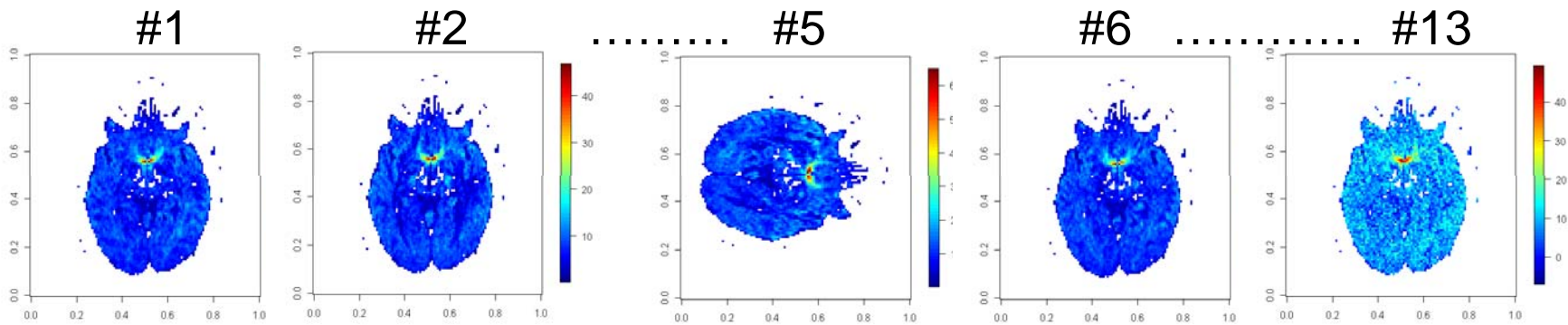
Uses Of The Symmetrical KL

How much more noisy is image A compared to image B?



Uses Of The Symmetrical KL

Detect outliers in a image series



Test similarity between #1 as
reference and rest based on
Brain mask of #1

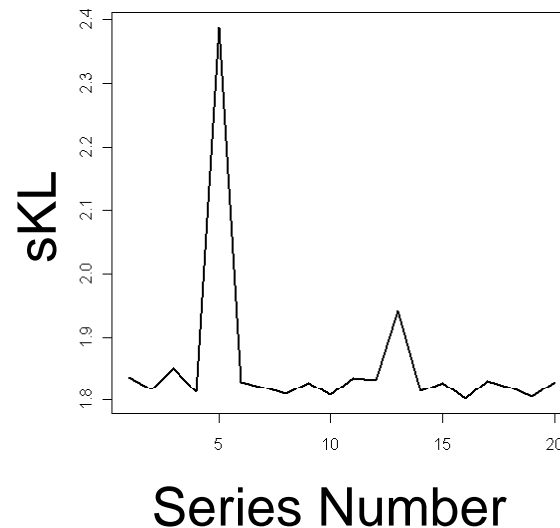


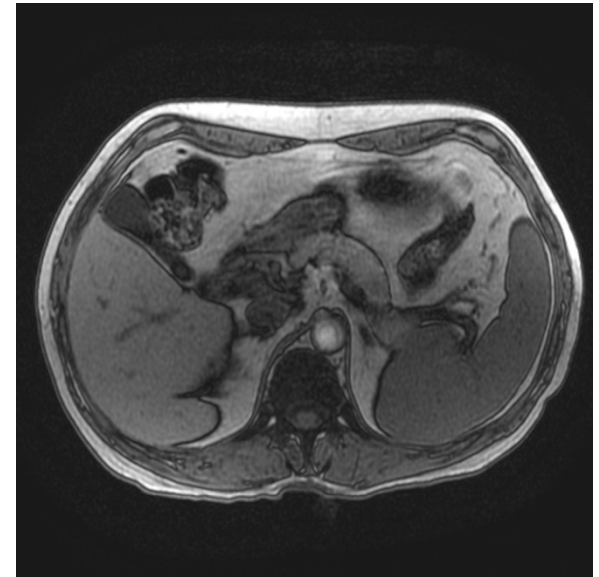
Image Feature Selection Using Information Theory

Objective

Learn how to use information theory for
defining imaging features

Mutual Information based Feature Selection Method

- We test the ability to separate two classes (features).
- Let X be the feature vector, e.g. co-occurrences of intensities
- Y is the classification, e.g. particular intensities
- How to maximize the separability of the classification?



Mutual Information based Feature Selection Method

- M.I. tests a feature's ability to separate two classes.
 - Based on definition 3) for M.I.

$$I(X, Y) = \sum_{x_i \in X} \sum_{y_i \in Y} p(x_i, y_i) \cdot \log \left(\frac{p(x_i, y_i)}{p(x_i)p(y_i)} \right)$$

- Here X is the feature vector and Y is the classification
 - Note that X is continuous while Y is discrete
- By maximizing the M.I. We maximize the separability of the feature
 - Note this method only tests each feature individually

Joint Mutual Information based Feature Selection Method

- Joint M.I. tests a feature's independence from all other features:

$$I(X_1, X_2, \dots, X_N; Y) = \sum_{k=1, N} I(X_k; Y \mid X_{k-1}, X_{k-2}, \dots, X_1)$$

- Two implementations proposed:
 - 1) Compute all individual M.I.s and sort from high to low
 - 2) Test the joint M.I of current feature while keeping all others
 - Keep the features with the lowest JMI (implies independence)
 - Implement by selecting features that maximize:

$$I(X_j, Y) - \beta \cdot \sum_k I(X_k, X_j)$$

Mutual Information Feature Selection Implementation Issue

- M.I tests are very sensitive to the number of bins used for the histograms
- Two methods used:
 - Fixed Bin Number (100)
 - Variable bin number based on Gaussianity of data

$$M_{bins} = \log N + 1 + \log(1 + \kappa \cdot \sqrt{N / 6})$$

- where N is the number of points and k is the Kurtosis

$$\kappa = \frac{1}{\sigma^4 \sqrt{24N}} \cdot \sum_{k=1, N} (x_k - \bar{x})^4 - \sqrt{\frac{3N}{8}}$$

Image Classification Based On Information Theoretic Measures

Objective

Learn how to use information theory for
imaging classification

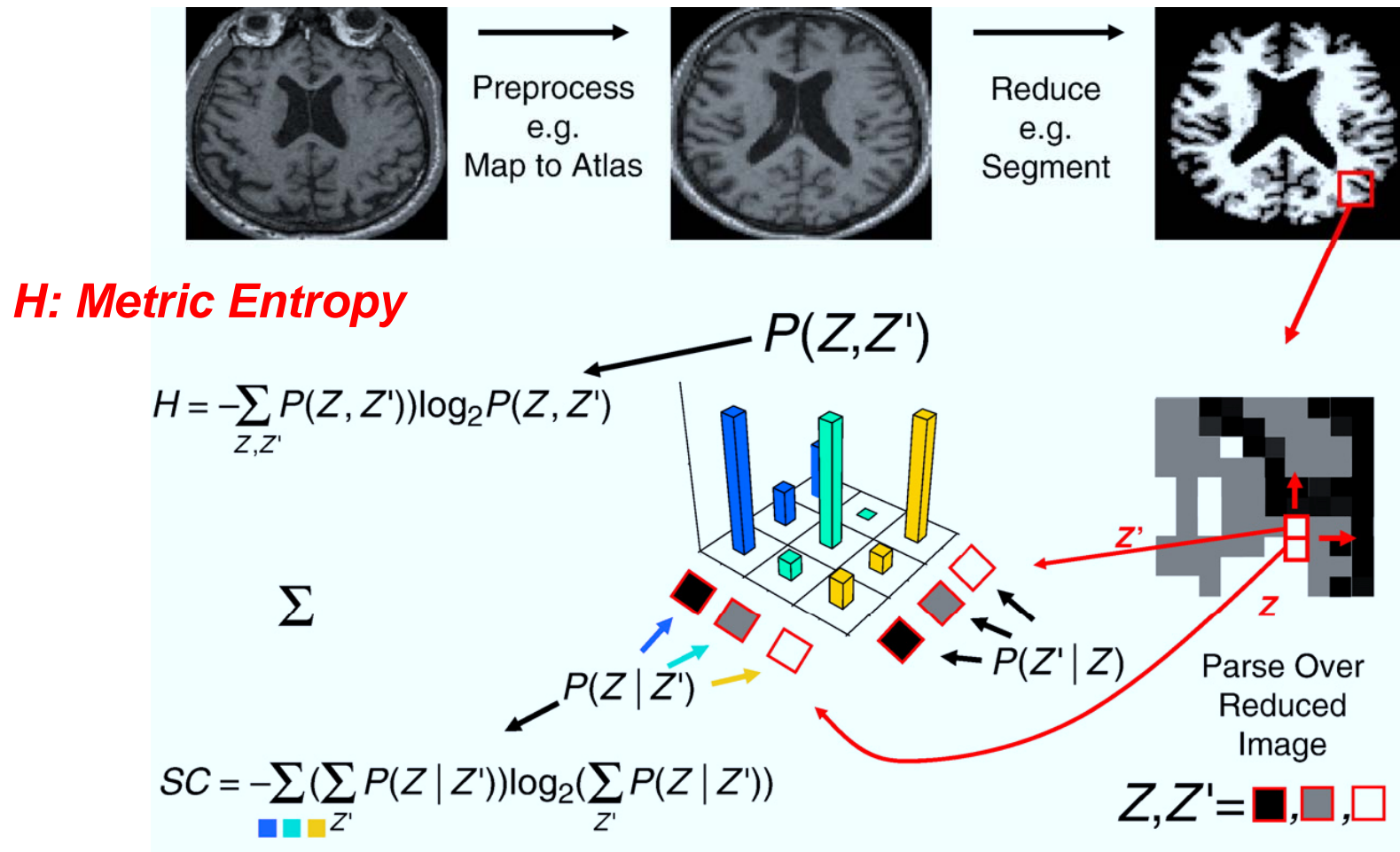
Learn how to quantify image complexity

Complexity In Medical Imaging

Complexity

- Many strongly interacting components introduce an inherent element of uncertainty into observation of a complex system
- In imaging, local correlations may introduce uncertainty in predicting intensity patterns
- Complexity describes the inherent difficulty in reducing uncertainty in predicting patterns
- Need a metric to quantify information, i.e. reduce uncertainty, in the distribution of observed patterns

Proposed Measures Of Complexity



SC: statistical complexity

By Karl Young

Proposed Complexity Measures

II

- **Metric Entropy (H)** – measures number and uniformity of distributions over observed patterns (joint entropy). For example, a higher H represents increasing spatial correlations in image regions
- **Statistical Complexity (SC)** – quantifies the information, i.e. uncertainty, contained in the distribution of observed patterns. For example, a higher SC represents an increase in locally correlated patterns.
- **Excess Entropy (EE)** – measures convergence rate of metric entropy. A higher EE represents increase long range correlations across regions

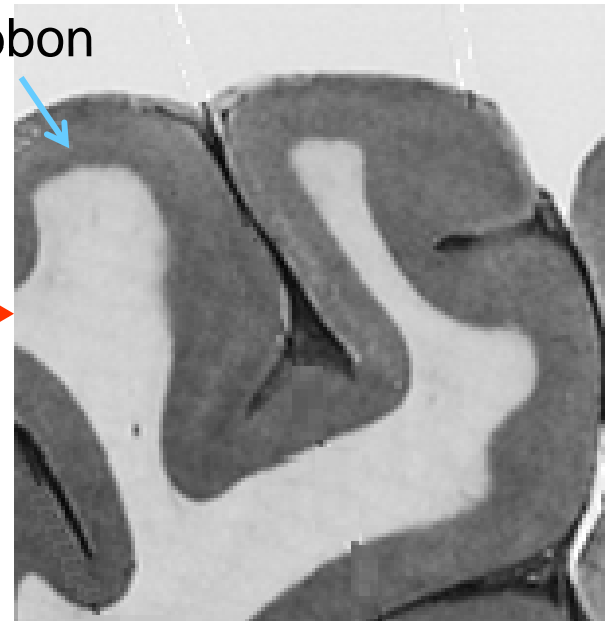
Application: Capturing Patterns Of Cortical Thinning

Brain MRI



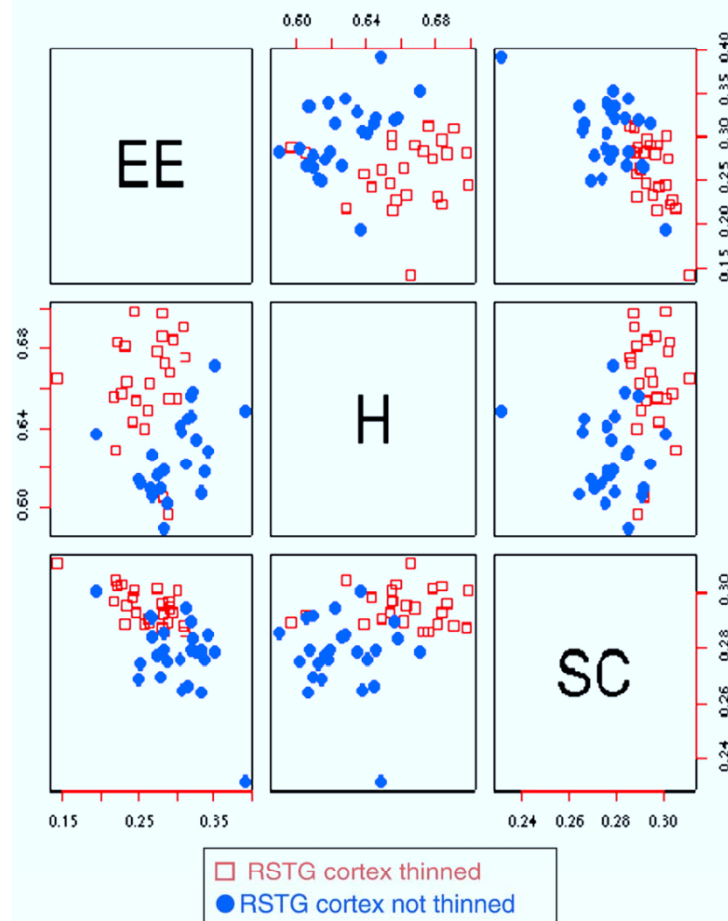
Histological Cut

Cortical ribbon

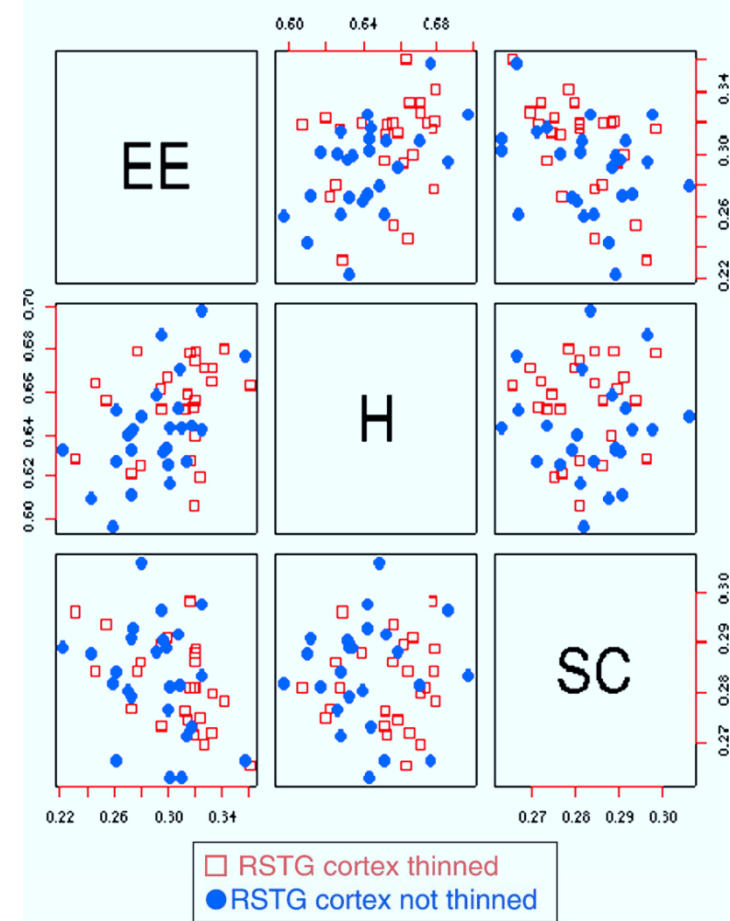


Complexity Based Detection Of Cortical Thinning (Simulations)

Right Superior Temporal Gyrus (Cortical Thinning in 25 Subjects)



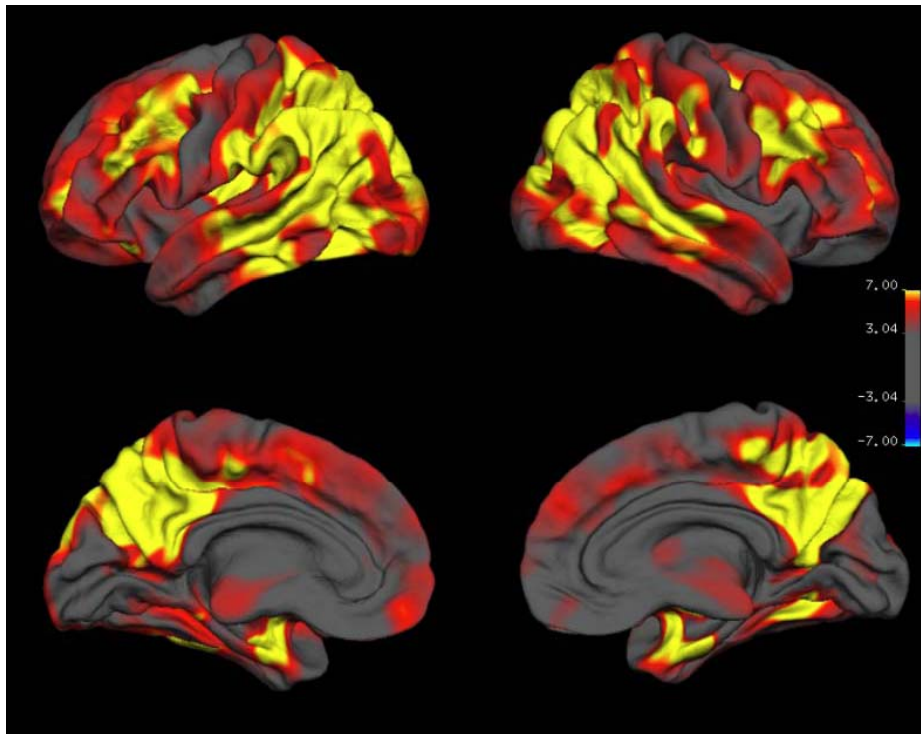
Left Superior Temporal Gyrus (No Cortical Thinning)



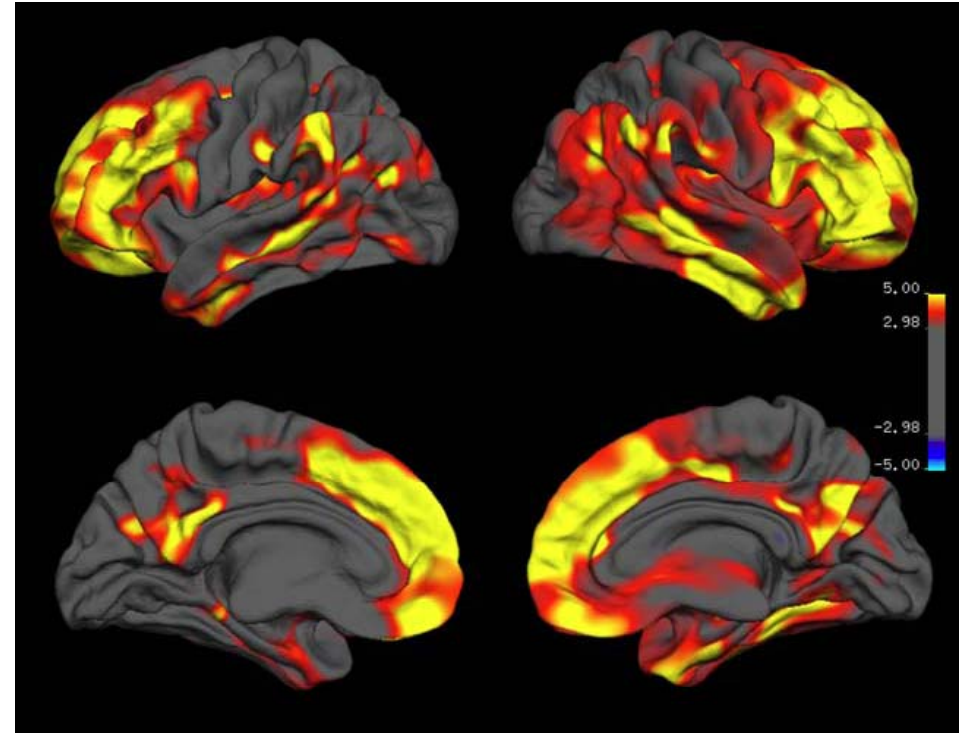
H, SC, and EE automatically identify cortical thinning and do a good job at separating the groups for which the cortex was thinned from the group for which there was no thinning

Patterns Of Cortical Thinning In Dementia Using Voxelbased Morphometry

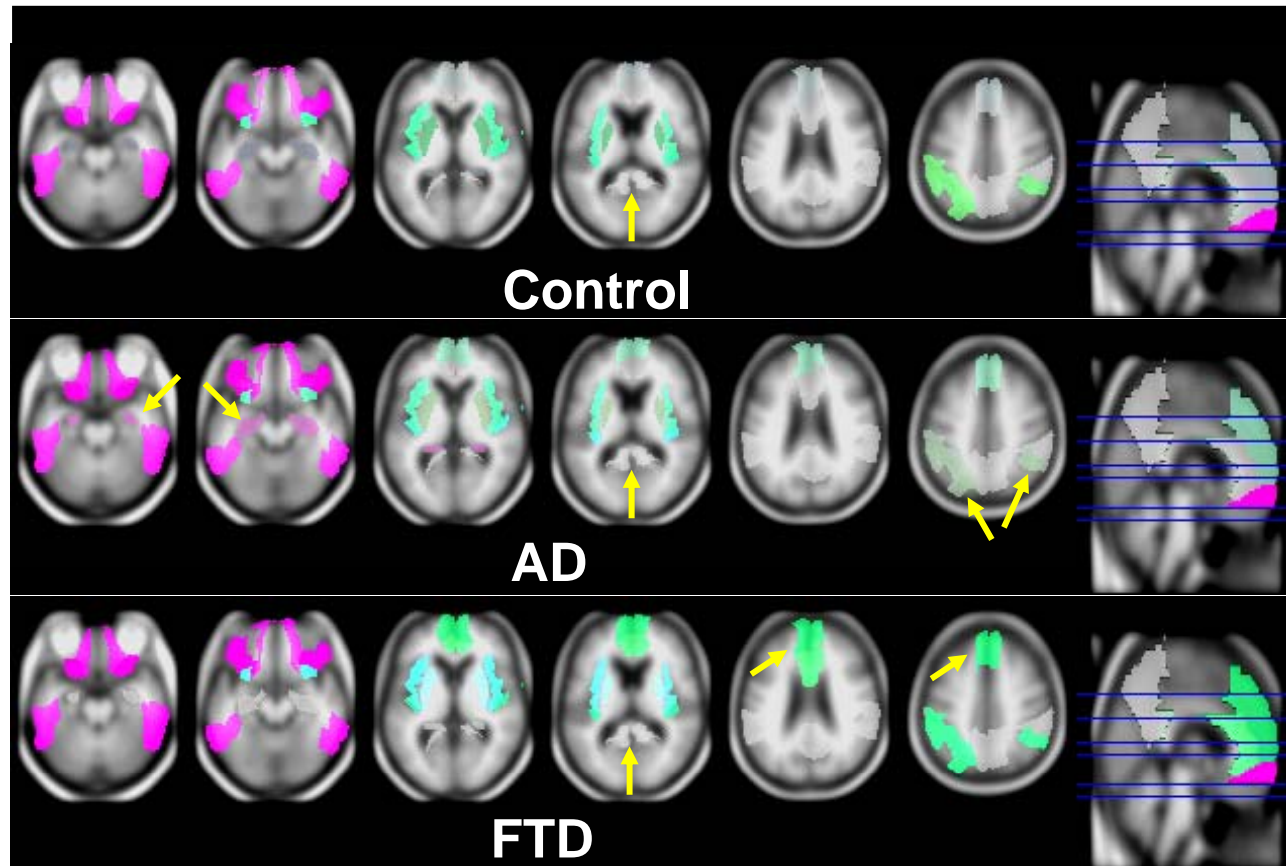
Alzheimer's Disease



Frontotemporal Dementia



Detection Of Cortical Thinning Pattern In Dementias Using Complexity Measures



RGB representation: **H**; **EE**; **SC**;;

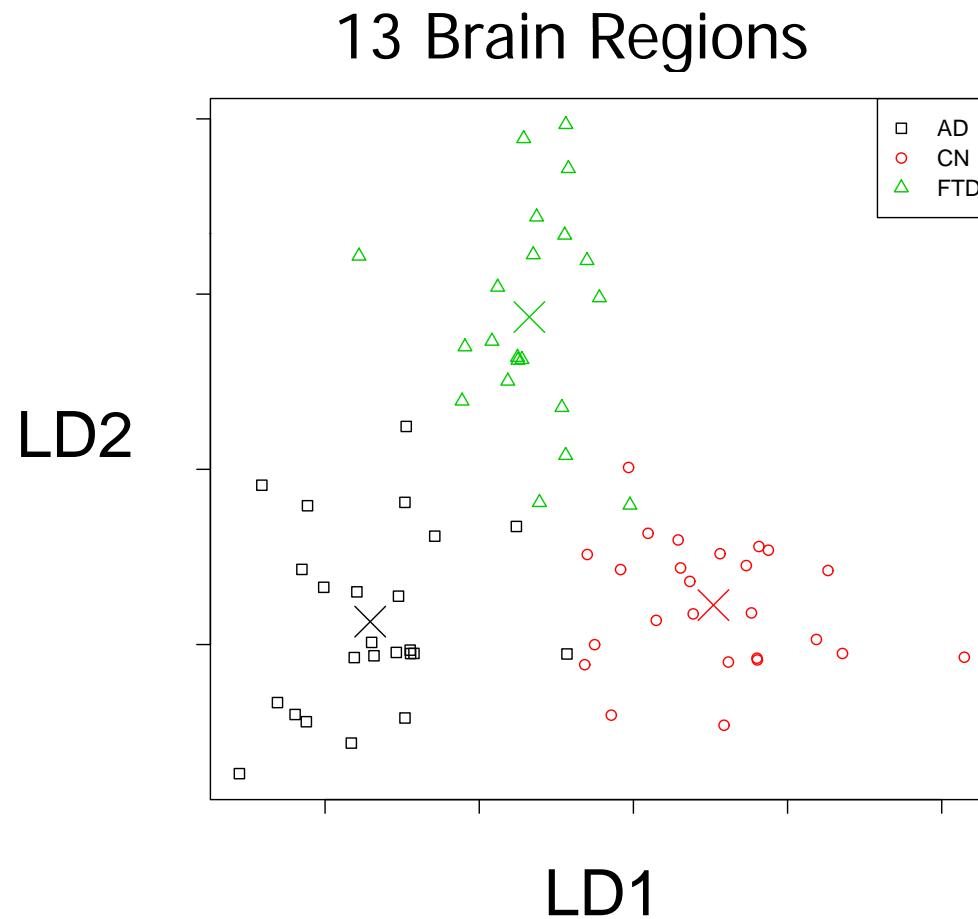
More saturated red means more spatial correlations;

More saturated green means more long range spatial correlations;

More saturation of blue means more locally correlated patterns;

Simultaneous increase/decrease of H,EE,SC results in brighter/darker levels of gray

Complexity Based Classification



Summary

Metric	Description
Entropy	Decrease in uncertainty of a random observation
Joint entropy	Simultaneous decrease in uncertainty of multiple random observations
Conditional entropy	Remaining uncertainty of one random observation given the probability of observations from another distribution
Mutual Information	Mutual dependence of random distributions
Kullback-Leibler divergence	Similarity between random distributions
Statistical complexity	Uncertainty in the distribution of correlated patterns
Excess entropy	Convergence of joint entropy

Literature

986

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 22, NO. 8, AUGUST 2003

Mutual-Information-Based Registration of Medical Images: A Survey

Josien P. W. Pluim*, *Member, IEEE*, J. B. Antoine Maintz, and Max A. Viergever, *Member, IEEE*